F-SQUARES, REPEATED MEASURES, AND NEAREST NEIGHBOR DESIGN

BU-914-MA September, 1986

Walter T. Federer
Biometrics Unit, Cornell University
Ithaca, New York, USA

In *Essays in Honor of F. A. Graybill*
J. N. Srivastava, Editor

Partially supported by the U.S. Army Research Office through the
Mathematical Sciences Institute of Cornell University.

F-SQUARES, REPEATED MEASURES, AND NEAREST NEIGHBOR DESIGNS

Walter T. Federer

Biometrics Unit
Cornell University
Ithaca, New York
U.S.A.

Two classes of nearest neighbor row-column designs are constructed using repeated measures and F-square theory. In the first class, any variety is bordered equally by all other varieties except itself. In the second class of designs, all varieties are bordered by all other varieties and also by itself. A competitive effects model is presented as an alternative to trend and correlation models currently in use for nearest neighbor designs.

INTRODUCTION

F-squares (FS) designs are a generalization of Latin square designs wherein some or all of the symbols appear more than once in a row and column. Repeated measures (RM) designs are special forms of Latin squares wherein all symbols follow and are followed by each other in columns (sequences). Nearest neighbor (NN) designs in blocks have each symbol appearing next to every other symbol and sometimes itself. In row-column nearest neighbor designs, every symbol appears next to every other symbol and also itself for some designs. Our purpose here is to demonstrate how to construct row-column NN designs using FS and RM designs and a competitive effects model is used to obtain a statistical analysis for the constructed designs. Note that trends could also be considered.

A CONSTRUCTION METHOD FOR NN DESIGNS

A method for constructing a RM design for v even symbols is the following. A cyclic Latin square of order v is put in standard order. Then, the first v/2 rows of the original Latin square are put into the odd rows of a new square. The last v/2 rows of the original square are put in the even rows in reverse ordering, i.e., row v/2 + 1st row of the original square is placed in the v^{th} row of the new square, the v/2 + 2nd original row is placed in the v - 2nd row of the new square, and so forth down to placing the v^{th} row of the original square in the 2nd row of the new square. To illustrate let v=6. Then,

Original square		New square = RM design
A B C D E F		A B C D E F
B C D E F A		F A B C D E
C D E F A B	→	B C D E F A
D E F A B C		E F A B C D
E F A B C D		C D E F A B
F A B C D E		D E F A B C

In RM designs obtained as above, every symbol follows and is followed by every other symbol except itself. Such a design is said to be balanced for one period residual effects. The rows represent periods and the columns are sequences of treatments on a sampling unit.

Construction Method I. Applying the above construction method for RM designs to *both* rows and columns of the original square results in a NN design in which every symbol follows and is followed by every other symbol one time in both rows and columns.

To illustrate, apply the RM method to the columns of the RM design obtained above resulting in the following:

A	F	B	E	C	D
F	E	A	D	B	C
B	A	C	F	D	E
E	D	F	C	A	B
C	B	D	A	E	F
D	C	E	B	F	A

Note that in the above design every symbol precedes and follows every other symbol once in rows and once in columns. Such a design may be used as a row-column NN design; it exists for all even n and for some odd numbers (but not 3, 5, or 7).

Construction Method II. This procedure consists of the following steps:

1. Construct a FS(2v;2^v) as the Kronecker product of a J matrix of side 2 (all ones) and a cyclic Latin square of order v, LS(v), i.e., J_2 * LS(v), v any integer (where * denotes the Kronecker product).
2. Apply the above RM procedure to the rows of the FS(2v;2^v).
3. Apply the above RM procedure to the columns of the result obtained in step 2.

The method is illustrated for v=3 and 4:

<u>Step 1</u>

F(6;2,2,2)

A	B	C	A	B	C
B	C	A	B	C	A
C	A	B	C	A	B
A	B	C	A	B	C
B	C	A	B	C	A
C	A	B	C	A	B

F(8;2,2,2,2)

A	B	C	D	A	B	C	D
B	C	D	A	B	C	D	A
C	D	A	B	C	D	A	B
D	A	B	C	D	A	B	C
A	B	C	D	A	B	C	D
B	C	D	A	B	C	D	A
C	D	A	B	C	D	A	B
D	A	B	C	D	A	B	C

<u>Step 2</u>

RM(6;2,2,2)

A	B	C	A	B	C
C	A	B	C	A	B
B	C	A	B	C	A
B	C	A	B	C	A
C	A	B	C	A	B
A	B	C	A	B	C

RM(8;2,2,2,2)

A	B	C	D	A	B	C	D
D	A	B	C	D	A	B	C
B	C	D	A	B	C	D	A
C	D	A	B	C	D	A	B
C	D	A	B	C	D	A	B
B	C	D	A	B	C	D	A
D	A	B	C	D	A	B	C
A	B	C	D	A	B	C	D

<u>Step 3</u>

NN(6;3;2,4)

A	C	B	B	C	A
C	B	A	A	B	C
B	A	C	C	A	B
B	A	C	C	A	B
C	B	A	A	B	C
A	C	B	B	C	A

NN(8;4;2,4)

A	D	B	C	C	B	D	A
D	C	A	B	B	A	C	D
B	A	C	D	D	C	A	B
C	B	D	A	A	D	B	C
C	B	D	A	A	D	B	C
B	A	C	D	D	C	A	B
D	C	A	B	B	A	C	D
A	D	B	C	C	B	D	A

In both rows and columns every symbol precedes and follows every other symbol four times. Every symbol precedes and follows itself twice in rows and in columns. In the above notation NN(2v; v; 2,4) 2v is the order of the square, there are v symbols; 2 denotes the number of times any symbol precedes and follows itself; and 4 denotes the number of times a symbol precedes and follows another symbol in both rows and columns.

The randomization procedure is to randomly allot treatments (varieties) to the symbols in the design. Note that the sequences are fixed in both rows and columns in order to retain the balanced arrangement for neighbors. If the experiment is at several sites, it may be desirable to use a new randomization of treatments to symbols at each site. If the sites are random sites and the experimental units are random variables, it may be unnecessary to rerandomize at each new site.

A RESPONSE MODEL FOR NN DESIGNS

Present response models for NN designs and the resulting statistical analyses are concerned with trends and correlations of neighbors. The following competitive effects model is offered as an alternative to the above. It is somewhat similar to a repeated measures model with carry-over effects. For small plots planted in the usual manner for agricultural experiments, it is believed that competition exists between neighbors. Whenever root systems of adjoining experimental units overlap, competition exists. Whenever one experimental unit is shaded by a second, competition exists. It is known for maize, for example, that experimental units which are closer than the height of the maize plants can be affected by competition since the root spread on one side is approximately one half the height of the plant.

Let the response model equation for the yield of variety h in the f^{th} row and g^{th} column be denoted as:

$$Y_{fghijkm} = \mu + \rho_f + \gamma_g + \tau_h + \alpha_{1i} + \alpha_{2j} + \alpha_{3k} + \alpha_{4m} + \varepsilon_{fghijkm} \qquad (1)$$

where μ is a general mean effect, ρ_f is the f^{th} row effect, γ_g is the g^{th} column effect, τ_h is the h^{th} variety effect, α_{1i}, α_{2j}, α_{3k}, and α_{4m} are the competitive effects of the four neighbors of variety h (diagonal competing effects are ignored), and the $\varepsilon_{fghijkm}$ are random error effects distributed with mean zero and common error variance σ^2_ε, $f,g,h = 1,\cdots,v$ and $i,j,k,m = 1,2,\cdots,v,x$, where x denotes border (blank or variety x surrounding the experiment). Note that the numbers 1, 2, 3, and 4 on the αs refer to the four neighbors of any experimental unit fgh.

Let $\mathbf{Y}_{v^2\times 1}$ be the vector of the v^2 yields in the experiment, X be the design matrix, and $\boldsymbol{\beta}_{4v+2}$ be the parameter vector. Then,

$$\mathbf{Y}_{v^2\times 1} = X\boldsymbol{\beta}_{4v+2} + \varepsilon_{v^2\times 1} \ , \qquad (2)$$

where $\varepsilon_{v^2\times 1}$ is the error vector. Using standard least squares theory, the normal equations are:

$$X'X\boldsymbol{\beta} = X'Y \quad . \qquad (3)$$

Since equation (3) is overparameterized, one may use the following constraints on the parameters to obtain solutions:

$$\sum_{f=1}^{v}\rho_f = \sum_{g=1}^{v}\gamma_g = \sum_{h=1}^{v}\tau_h = \sum_{i=1}^{v}\alpha_i = 0 \quad , \qquad (4)$$

where the position of a neighbor is ignored in the αs. An analysis of variance breakdown of the v^2 degrees of freedom is presented in Table 1.

To illustrate how to construct the X matrix in (2) for v=6, consider the responses for fgh = 11A and 32A for the design illustrated for Construction Method I:

$$Y_{11AXXFF} = \mu + \rho_1 + \gamma_1 + \tau_A + 2\alpha_X + 2\alpha_F + \varepsilon_{11AXXFF}$$

$$Y_{32ABCDE} = \mu + \rho_3 + \gamma_2 + \tau_A + \alpha_B + \alpha_C + \alpha_D + \alpha_E + \varepsilon_{32ABCDE} \ .$$

With the constraints in (4), the various totals in terms of parameters for the design using Construction Method I are:

$$Y_{\cdot\cdot\cdot\cdot\cdot\cdot\cdot} = v^2\mu + 4v\alpha_x$$

$$Y_{1\cdot\cdot\cdot\cdot\cdot\cdot} = v(\mu + \rho_1) + 6\alpha_x - \text{(the } 2\alpha\text{s for varieties at the ends of row one)}$$

$$Y_{2\cdot\cdot\cdot\cdot\cdot\cdot} = v(\mu + \rho_2) + 2\alpha_x - \text{(the } 2\alpha\text{s for varieties at the ends of row two)}$$

$$Y_{\cdot g\cdot\cdot\cdot\cdot\cdot} = v(\mu + \gamma_g) + \text{(a similar structure for rows because design is symmetric)}$$

$$Y_{\cdot\cdot h\cdot\cdot\cdot\cdot} = v(\mu + \tau_h) - 4\alpha_x - 4\alpha_h \ .$$

These totals for the αs are obtained from $X'Y$. Likewise, the variances of any estimable contrast among the parameters may be obtained from:

$$(X'X - R)^{-1}\sigma^2_\varepsilon \ , \tag{5}$$

where R is a matrix which subtracts (4) from $X'X$. A similar approach leads to a statistical analysis for designs obtained from Construction Method II.

The above analyses were developed assuming equal "neighboring," i.e., each neighbor bordered a given experimental unit by the same amount. If this is not true, it is suggested that proportionate weights be given to the αs. For example, each α was given a weight of one for the above. If an experimental unit (e.u.) is ℓ units long and w units wide, then two neighbors border the e.u. ℓ units and the other two only w units. In equation (1) it is suggested that weights $4\ell/(2\ell+2w) = 2\ell/(\ell+w)$ be given to the two neighbors with borders equal to ℓ and weights $2w/(\ell+w)$ be given to the two neighbors whose borders are of length w. The αs from this procedure would be on the same basis then as the αs considered above.

Table 1. Partitioning of degrees of freedom using response model equation (1)

Source of variation	Method I	Method II
Total	v^2	$4v^2$
Correction for mean	1	1
Rows (ignoring competition)	v-1	2v-1
Columns (ignoring competition)	v-1	2v-1
Varieties (ignoring competition)	v-1	v-1
Border effect (eliminating rows, columns, and varieties)	1	1
Competing effects (eliminating all other effects)	v-1	v-1
Error	v^2-4v+2	$4v^2-6v+2$